EAST Search History

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S1	114	715/542	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 12:50
S3	3882	(structure or format$4) SAME extract$3 SAME character$1	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 08:56
S4	2836	S3 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 08:57
S5	242	S4 and (font and color and size)	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 08:57
S6	8	S5 and (typographic)	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 12:55
S7	3882	(structure or format$4) SAME extract$3 SAME character$1	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 15:43
S8	2836	S7 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 15:46
S9	37	S8 and (font and (bold or bold$face) and color and size)	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 14:49
S10	3	("5764241" | "6442523" | "6819394").PN. OR ("7071942").URPN.	US-PGPUB; USPAT; USOCR	OR	ON	2006/11/29 14:13
S11	26	("3233219" | "4769716" | "5303313" | "5321773" | "5438630" | "5440651" | "5469512" | "5526444" | "5542006" | "5566247" | "5577166").PN. OR ("5706364").URPN.	US-PGPUB; USPAT; USOCR	OR	ON	2006/11/29 14:35
S12	22	S11 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 14:38

S13	2	("20060143555").PN.	US-PGPUB; USPAT; JPO; DERWENT	OR	OFF	2006/11/29 14:38
S14	0	("2006/0143555").URPN.	USPAT	OR	ON	2006/11/29 14:39
S15	4842	document SAME extract$3 SAME character$1	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 14:48
S16	3446	S15 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 14:48
S17	64	S16 and (font and (bold or bold$face) and color and size)	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 15:49
S18	64	S17 and extract$4	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 14:50
S19	4	S18 and typographic$4	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 14:50
S20	26	("3233219" | "4769716" | "5303313" | "5321773" | "5438630" | "5440651" | "5469512" | "5526444" | "5542006" | "5566247" | "5577166").PN. OR ("5706364"). URPN.	US-PGPUB; USPAT; USOCR	OR	ON	2006/11/29 14:52
S21	572	(structure or format$4 OR template) AND extract$3 AND character$1 AND typographic$4	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 17:29
S22	380	S21 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 17:30
S23	28	S22 and (font and (bold or bold$face) and color and size)	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 17:30
S24	42	("4599692" | "5020112" | "5237627" | "5303313" | "5321773" | "5333275" | "5526444").PN. OR ("5594809").URPN.	US-PGPUB; USPAT; USOCR	OR	ON	2006/11/29 17:15

S25	40	S24 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 17:32
S26	739	(structure or format$4 OR template) AND (extract$3 or pars$3) AND character$1 AND typographic$4	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 06:32
S27	476	S26 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 06:33
S28	42	S27 and (font and (bold or bold$face) and color and size)	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 06:34
S29	6	(structure or format$4 OR template) SAME (extract$3 or pars$3) SAME character$1 SAME typographic$4	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 17:31
S30	4	S29 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/29 17:32
S31	20	(extract$4 or pars$4) SAME character$1 SAME typographic$4	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 06:33
S32	10	S31 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 06:58
S33	0	S32 and (font and (bold or bold$face) and color and size)	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 06:56
S34	22289	(pars$3 or extract$3) with (character$1 or (character ADJ string$))	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 06:55
S35	227	S34 and (font and (bold or bold$face) and color and size)	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 06:58
S36	96	S35 and HTML	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 06:58

12/6/2006 9:45:27 AM

Page 3

C:\Documents and Settings\jdebrow\My Documents\EAST\Workspaces\10768178 extracting information frm a formatted document.wsp

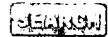| S37	53	S36 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 07:21
S38	30	("5721827" | "5778368" | "5832494" | "5881131").PN. OR ("6466940"). URPN.	US-PGPUB; USPAT; USOCR	OR	ON	2006/11/30 07:20
S39	28	S38 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 07:48
S40	47	("20010009017" | "20030221163" | "20030225763" | "4270182" | "5319777" | "5764906" | "5813006" | "5835905" | "5895470" | "5918236" | "5923850" | "5924090" | "6052693" | "6065016" | "6076088" | "6094653" | "6112203" | "6122647" | "6128613" | "6212552" | "6253198" | "6260033" | "6266664" | "6269369" | "6301614" | "6314409" | "6336108" | "6336139" | "6349309" | "6377936" | "6389436" | "6418432" | "6463430" | "6466940" | "6493703" | "6529891" | "6553364" | "6556964" | "6601026" | "6618717" | "6640224" | "6654768" | "6668256" | "6675162" | "6697793" | "6745161" | "6859797").PN. OR ("7065483"). URPN.	US-PGPUB; USPAT; USOCR	OR	ON	2006/11/30 07:47
S41	45	S40 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 07:48
S42	143	tag$1 SAME extract$3 SAME character$1 SAME HTML	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 09:28
S43	63	S42 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 09:29
S44	10	S43 and (font and color and size)	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 09:29
S45	4	("2002/0065814").URPN.	USPAT	OR	ON	2006/11/30 09:23

S46	250	tag$1 SAME (extract$3 or pars$4) SAME character$1 SAME HTML	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 09:28
S47	129	S46 and (@ad<"20010803" or @rlad<"20010803")	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 09:29
S48	35	S47 and (font and color and size)	US-PGPUB; USPAT; JPO; DERWENT	OR	ON	2006/11/30 09:29

PØRTAL

USPTO

Search: ⊙ The ACM Digital Library   ○ The Guide

tag data extraction   [SEARCH]

THE ACM DIGITAL LIBRARY

ᵢ€ Feedback Report a problem Satisfaction survey

Terms used **tag** **data** **extraction**   Found **41,263** of **193,448**

Sort results by	relevance ▼	🔖 Save results to a Binder	Try an Advanced Search
Display results	expanded form ▼	⁇ Search Tips	Try this search in The ACM Guide
		☐ Open results in a new window	

Results 1 - 20 of 200   Result page: **1** 2 3 4 5 6 7 8 9 10   next
Best 200 shown   Relevance scale ☐▣▣▣■

1 Data extraction: Web data extraction based on partial tree alignment   ■
Yanhong Zhai, Bing Liu
May 2005 **Proceedings of the 14th international conference on World Wide Web**
Publisher: ACM Press
Full text available: 🗎 pdf(357.50 KB)   Additional Information: full citation, abstract, references, index terms

> This paper studies the problem of extracting data from a Web page that contains several structured data records. The objective is to segment these data records, extract data items/fields from them and put the data in a database table. This problem has been studied by several researchers. However, existing methods still have some serious limitations. The first class of methods is based on machine learning, which requires human labeling of many examples from each Web site that one is interested in ...

> **Keywords**: data extraction, data record extraction, wrapper

2 Data extraction: Fully automatic wrapper generation for search engines   ■
Hongkun Zhao, Weiyi Meng, Zonghuan Wu, Vijay Raghavan, Clement Yu
May 2005 **Proceedings of the 14th international conference on World Wide Web**
Publisher: ACM Press
Full text available: 🗎 pdf(315.59 KB)   Additional Information: full citation, abstract, references, index terms

> When a query is submitted to a search engine, the search engine returns a dynamically generated result page containing the result records, each of which usually consists of a link to and/or snippet of a retrieved Web page. In addition, such a result page often also contains information irrelevant to the query, such as information related to the hosting site of the search engine and advertisements. In this paper, we present a technique for automatically producing wrappers that can be used to extr ...

> **Keywords**: information extraction, search engine, wrapper generation

3 Research sessions: Research 28: Search applications: Automatic extraction of   ■
dynamic record sections from search engine result pages
Hongkun Zhao, Weiyi Meng, Clement Yu
September 2006 **Proceedings of the 32nd international conference on Very large data bases - Volume 32 VLDB'2006**
Publisher: VLDB Endowment.